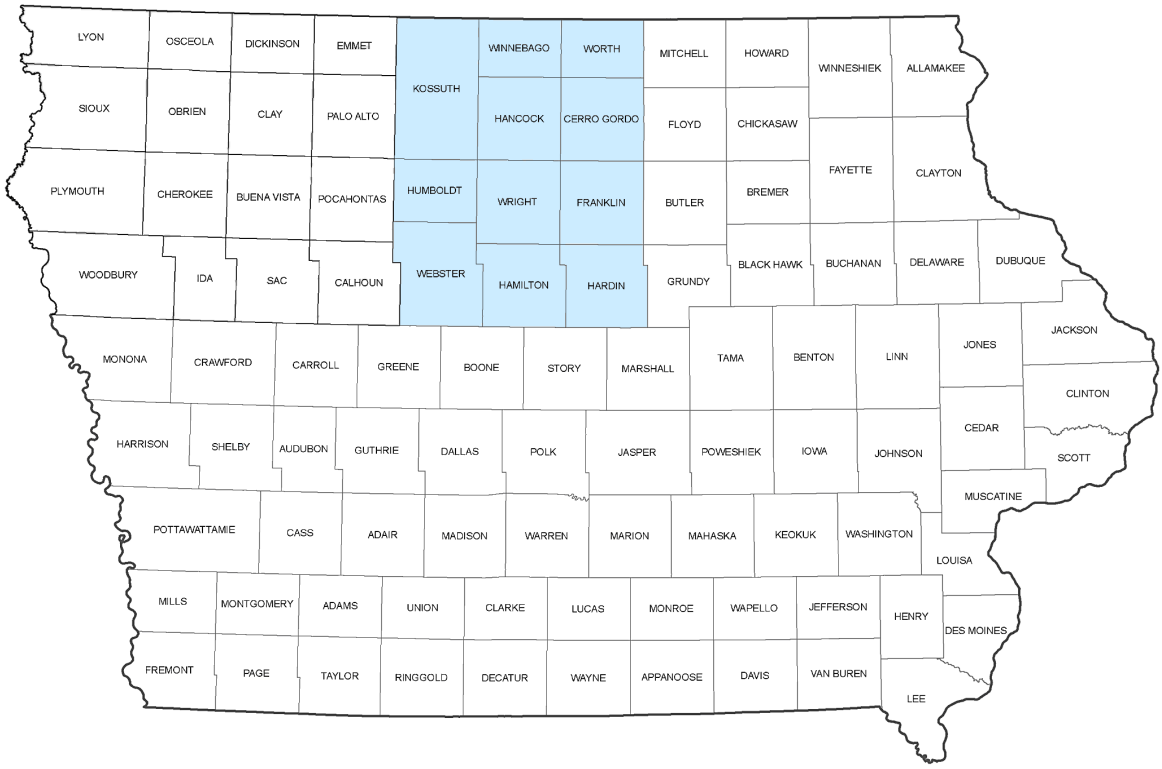
Digital Infrastructure and Metadata for Geologic Collections, NC Iowa



IGWS Contract Report
National Geological and Geophysical Data Preservation Program
United States Geological Survey Award Number G11AP20199

Iowa Geological & Water Survey
Iowa Department of Natural Resources

The project described in this publication was supported by award number G11AP20199 from the United States Geological Survey. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the USGS.

Introduction

This report documents the development of National Geological and Geophysical Data Preservation Program-standard metadata for the physical and electronic data collections maintained by the Iowa Geological and Water Survey (IGWS), and the conversion of data held as paper records to electronic formats for counties in southwest Iowa. This work was conducted as part of the U.S. Geological Survey - National Geological and Geophysical Data Preservation Program (NGGDPP), under award number G11AP20199 for 2011.

IGWS and Geologic Data Storage

The IGWS was established as the Iowa Geological Survey (IGS) in 1892. After periods of being part of the University of Iowa and later an independent state agency, IGS became part of the newly formed Iowa Department of Natural Resources in 1986. In 2008, the survey changed its name to Iowa Geological and Water Survey to reflect the agency's long-term role as the state's lead earth and water science agency.

Figure 1. Core and chip sample storage area in the IGWS Oakdale Research Facility and Rock Library building.

IGWS occupies two buildings on the University of Iowa Campus. Most staff offices are located on the main campus of the University in Trowbridge Hall, which is shared with the Department of Geosciences. The research facility and rock library building is located on the University's satellite Oakdale Campus. The 11,000-square foot Oakdale facility, built in 1983 then expanded to 12,250 square feet in 2000, functions as the repository for

most physical geologic material storage, such as cores, cuttings, and hand samples (fig. 1). Sample preparation, cataloging, and descriptions as well as a variety of laboratory work are carried out at the Oakdale building. The Oakdale facility provides storage space for the bulk of the paper files that document the physical collections. In addition a wide variety of older paper files and unpublished and archived maps are stored at the Oakdale facility.

Most critical and widely used geologic information for the state, mainly from cores and well cuttings, logs, and construction data, is electronically held in the IGWS Geologic Sample Database (GEOSAM) as a relational database developed to reflect the site-based nature of the data linked to a collection of electronic documents. Access is free and available to the general public and is provided through a web-based interface (http://www.igsb.uiowa.edu/webapps/geosam), which also provides tools to query the database and view the electronic documents. GEOSAM and GEOSAM-compatible databases form the backbone of the current IGWS data preservation system and its ongoing preservation plan.

Summary of 2011 Grant Activities

IGWS included the following in its 2011 NGGDPP grant proposal:

- Develop metadata, in accordance with National Catalog standards, for all current GEOSAM sites that can be located and fall within Iowa.
- Establish locations as needed for physical and paper data collections from 11 counties in north central (NC) Iowa (fig. 2) that were not previously held in GEOSAM.
- Make any needed changes to GEOSAM to accommodate different site/data types.
- Move these locations and site types to GEOSAM, develop metadata for these additions.
- Submit all developed metadata to the National Catalog.

The data collections planned for addition to GEOSAM from NC Iowa included:

- Outcrop records/descriptions from publications, field notes, and other sources.
- Physically archived hand specimens and bagged quaternary specimens.
- Mechanical and clay mineralogical data for Quaternary materials.
- Groundwater Quality Analyses.
- Additional information on existing well/cuttings/core records.

Initial examination of some of these collections identified the need for minor changes to the original plans. Little in the way of “mechanical or clay-mineralogical data for Quaternary materials” was found to be archived for this part of the state. The same is true for bagged / hand specimens. Thus, efforts were directed towards the preservation of spatial/map data and enhancing the collection of digital well log data.

We further proposed to continue scanning and making web-available the IGWS Annual Report series and related publications, with metadata provided to the National Catalog; and to continue cataloging and preservation of historic spatial/map data. The numbered volumes of the IGWS Annual Report series that have been scanned are currently available on line through http://ir.uiowa.edu/igsar/.

Significant data from outside of the NC Iowa area was also preserved this year. The reasons for this are two-fold. First, current IGWS projects and interests in Northeast Iowa required the digitization of some data in that area. Second, a variety of data sets and spatial data aren't amenable to a regional approach. As it was more time-efficient to process these materials together, we did so. Totals reported below for this year's project are broken into NC Iowa and statewide addition.

Summary of Project Results

NGGDPP Metadata Development

The IGWS collections developed for NGGDPP are broadly defined and may include physical samples, derived data, and documents. Iowa's data was prepared in XML format as the bulk of the data from which metadata was extracted is stored in SQL Server databases. XML was selected over a delimited text format as the simpler of the formats to develop from a relational database. The necessary relationships between tables and views are modeled as views/queries for output to the required format from this source as described below.

Metadata preparation was carried out as follows:

1) SQL server tables, views, and custom functions to export the necessary text strings were developed.
2) Developed queries to select the data from the views prepared in step one to a temporary XML-formatted data item then exported this data item to a text-based XML file
3) When the text-based XML was successfully exported it was validated against the NGGDPPMetadata.xsd schema using XML Notepad.
4) When the local validation was successful, the file was uploaded to the NGGDPP site and an email request was made to the database manager to replace or append the uploaded data

The NGGDPP metadata elements, along with a description of the data items provided for each element, are listed in table 1. In instances where required data elements were not available, metadata records were not prepared, e.g. if no additional identifier were stored for a data item, the alternateTitle item was omitted from the metadata.

The IGWS metadata supplies data for the optional items in some cases. For most collections, optional items were either populated for all records or no records. For example, the optional item, verticalExtent, is populated for most collections but was omitted for the Geochemical Samples collection (water quality) as not relevant to the data

being presented. In other cases, the optional items were populated where the data to be supplied was available.

Table 1. Description of metadata elements developed from IGWS data. Bolded elements indicate required data.

Tag/Data Element	Description of data supplied
collectionID	Collection ID from previously defined Iowa NGGDPP collections
title	Name of collection and IGWS identifier for the sample/log
alternateTitle	Other identifiers for the data item
abstract	General descriptive information about the site and/or data item including identifier and site type; other information, depending on the collection being documented, includes total depth, sample depth, completion date, collection date, aquifer, and author.
dataType	NGGDPP catalog supplied values
supplementalInformation	Statement concerning archive location of samples/logs and a URL where additional information may be accessed
coordinates	Longitude and latitude in decimal degrees (NAD83) derived from coordinates stored in Geologic Sample Database (GEOSAM); in the case of maps bounding coordinates were provided; otherwise a single x,y coordinate was provided
alternateGeometry	County, tier, range, section, and quarter sections for sites where this information is available; for map data the centroid x,y were provided
onlineResource/resourceURL	URL provided where applicable
browseGraphic/resourceURL	not used
dates/date	Last update of the record for some collections; sampling date for some collections
datasetReferenceDate	The date the metadata for the collection was prepared
verticalExtent	Total depth supplied for wells, length of section measured/described for field notes; omitted for other data types

Additions to IGWS Collections for NC Iowa and statewide

Eleven counties in NC Iowa (fig. 2) were selected for more focused development of digital data from samples not yet completely cataloged in IGWS databases or from archived documents. Additionally, large format maps and other geologic documents were scanned, and archived drillers logs were scanned and entered into GEOSAM. Documents that were scanned and cataloged from the focus area and other preservation projects completed in 2011 include the following:

- No field notes/measured section descriptions were entered for NC Iowa this year (78 had previously been entered), but 127 sets for other areas of Iowa were scanned and entered into GEOSAM.
- 4,761 well file folders from NC Iowa were examined, with relative information from 3,700 (Fig. 2) scanned and attached as PDF's to the existing well data in GEOSAM.
- A total of 1,122 large format maps were classified into six map types (Table 2) and scanned. Of these, 84 were georeferenced and 84 cataloged.
- 250 old, archived drillers logs were scanned and entered into the GEOSAM database
- In conjunction with the University of Iowa Libraries 38 "early series" IGS Annual Reports where scanned and are online at: http://ir.uiowa.edu/igsar/.

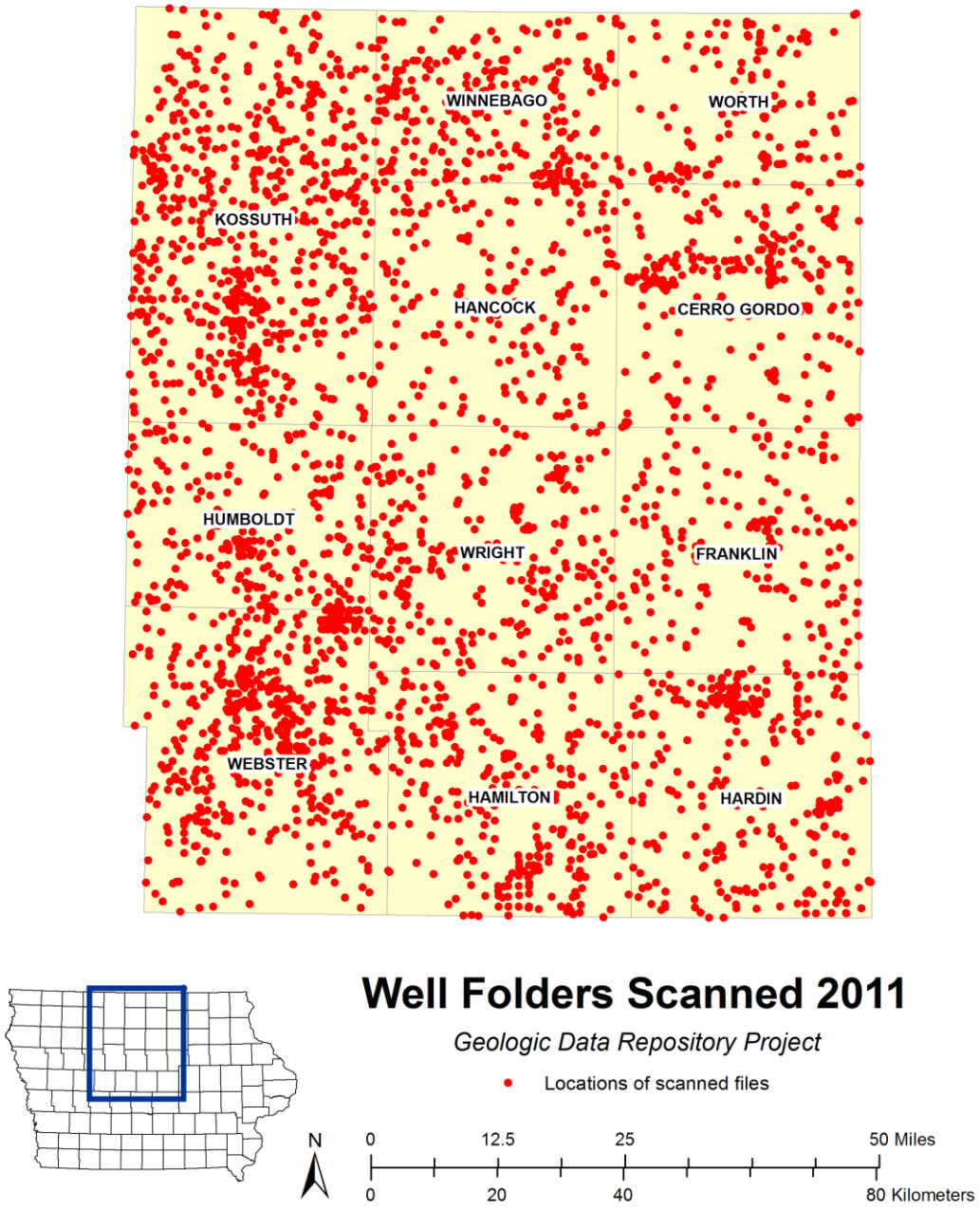


Figure 2. Map of NC Iowa study area showing the location of 3,700 well sites (of 4,761 well folders examined) with relevant information scanned and added to GEOSAM.

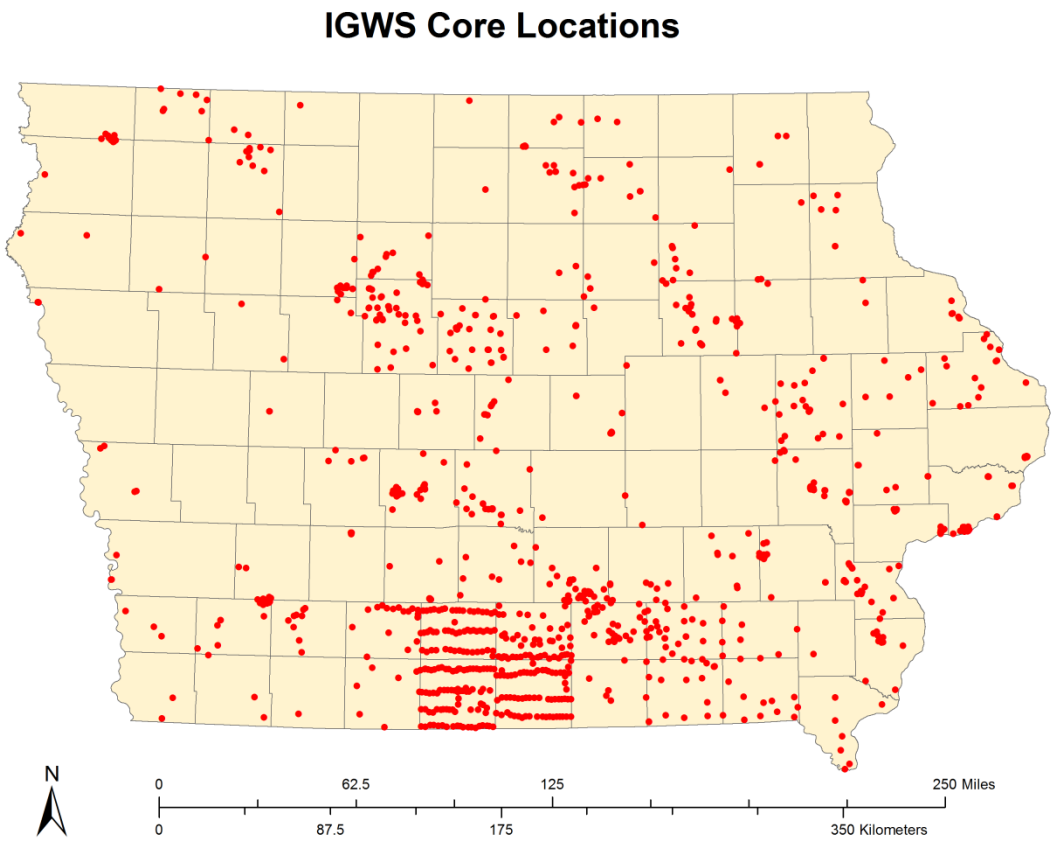


Figure 3. Map of Iowa showing the locations of the drilling sites of 1,220 drill cores that have been inventoried and updated in GEOSAM during 2011 as a part of this National Geological and Geophysical Data Preservation grant.

Inventory of IGWS Cores

IGWS' Oakdale facility houses approximately 450,000 feet of rock core samples, collected from 1,613 sites around Iowa. These cores, most donated to the IGWS, were originally drilled as a part of gas storage investigation and production, environmental engineering investigations, quarry and IGWS exploration activities, and for other reasons. The core sets are systematically stored (see Fig. 1) and cataloged for easy retrieval. Many of these cores have been logged by IGWS geologists, and the cores are frequently accessed for study and sampling by mineral industry exploration geologists, for research and as teaching tools by geologists and students from various academic institutions, and by IGWS geologists in the course of their geologic investigations.

Organization of these core samples is critical to utilizing this resource efficiently. As such, the entire core library was inventoried for purposes of ensuring electronic cataloging quality. Inventorying consisted of recording sample footages, box numbers, and storage locations for all of 1,613 shelved core sets. Of these, 1,220 had sufficient information for inclusion in the inventory, the remaining 393 will need additional research to recover missing information. The inventory prompted several updates to the sample storage cataloging tables in the GEOSAM database. In addition, core sets that

had been previously described and assigned stratigraphic units were entered into the core stratigraphy tables in GEOSAM tables. The availability of this information as an electronic resource makes searching the core sample collection by stratigraphic unit possible, greatly improving the collection as an information resource.

Archived Drillers' Logs Scanned and entered in GEOSAM

In the course of organizing archived material at the IGWS over the past few years, a number of boxes of drillers' logs were discovered. Initial investigation disclosed that these logs had never been entered into the IGWS GEOSAM database. As a part of this National Geological and Geophysical Data Preservation grant, approximately 250 driller's logs were assigned locations and entered into GEOSAM during the project year. Additionally, the logs were scanned and uploaded to the document collection linked to GEOSAM. These logs contain valuable geologic, hydrogeologic, casing, and location data that is now accessible electronically.

Table 2. Large-format documents including maps, cross-sections, stratigraphic columns, etc classified, scanned, and cataloged in fy 2011.

Document Type	Number Scanned	Number of maps Georeferenced	Number Added to Catalog
Geologic	953	64	64
Structure	5	1	1
Thickness	13	13	13
Hydrogeologic	8	2	2
Geophysics	40	3	3
Other	113	1	1
Totals	**1,122**	**84**	**84**

Preparing Map Documents for Data Preservation

IGWS has a large collection of paper maps and other documents that exceed page-size (8-1/2 x 11) dimensions that have been prepared during its 120 year history. Most of these documents were never intended for publication, but were used as working maps or reference maps by the IGWS geologists. Much of the content of the maps has been superseded by more recent work; however, the earlier maps are viewed as work that provided material contributions to currently accepted interpretations of a variety of geologic problems. Preservation of these maps and other documents is considered a key task for this reason.

During FY11 1,122 maps (Table 2) and other oversize documents (larger than 8½ x 11 in.) were prepared for preservation as digital images. Initially, the maps were reviewed by a staff geologist, assigned a unique identifier, and assigned to one of six thematic categories based on the content of the map. A significant addition to the map collection during the FY2011 project year was IGWS's collection of coal mine maps.

Following the initial organizing and cataloging steps, the maps and related documents were scanned using a Contex wide-format scanner. The files were saved as uncompressed .tiff images as either 24-bit (full color) or gray scale images depending on the original document. During scanning, the authors' names, dates, scanning resolution, bit depth, were captured in a spreadsheet that was used to organize the scanning and georeferencing phases. Minor edits were performed on the scanned images including straightening and cropping. Map document that encompassed all or part of Iowa were georeferenced to the UTM NAD83, Zone 15 coordinate system which has been adopted as a standard by Iowa DNR. All of the georeferenced images were loaded into ESRI raster catalogs and attribute tables populated with basic identifying information, and map centroids and bounding coordinates calculated and added to the attributes. All data were exported to a MS SQL Server database that was used to prepare the metadata.

Metadata Records added to the National Catalog

Metadata describing the data and physical collections were loaded into the National Catalog in December, 2012. This number of data items included in the metadata are summarized by collection as follows:

Records for NC Iowa	**fy 2011 entries**	**previous entries**
Rock Cores	0 records	171 records
Well Cutting Samples	15 records	4,412 records
Well Logs (Strip & Drillers)	1,206 records	11,914 records
Scanned Maps & Documents (non-coal)	0 records	0 records
Scanned Coal Mine Maps	0 records	0 records
Field Notes and Sections	0 records	78 records

Records for Other Iowa Regions	**fy 2011 entries**	**Iowa totals**
Rock Cores	480 records	1,672 records
Well Cutting Samples	4,892 records	480 records
Well Logs (Strip & Drillers)	4,891 records	89,255 records
Scanned Maps & Documents (non-coal)	1,123 records	2,249 records*
Scanned Coal Mine Maps	0 records	1,515 records**
Field Notes and Sections	127 records	4,174 records

* total georeferenced 923 records
(georeferenced in 2011 197 records)

** total georeferenced (2011) 243 records

Award 2011 Summary and Award 2012 Plans

The goals proposed for the 2011 project year were met with the minor changes described above. Updated metadata was created for cores, cuttings, well logs, field notes, and maps collection types described for Iowa. No significant changes occurred within other previously submitted collections types.

For the IGWS 2012 award from NGGDPP, the focus will be on developing digital data from paper archives will shift to a block of counties in northeast Iowa. Updated metadata will be developed near the end of the project period to replace the records provided in 2011. Work to improve integration of the new digital data types into GEOSAM will continue as needed.